

Probability Review

Gonzalo Mateos

Dept. of ECE and Goergen Institute for Data Science

University of Rochester

`gmateosb@ece.rochester.edu`

`http://www.ece.rochester.edu/~gmateosb/`

September 7, 2017

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ An event is something that happens
- ▶ A random event has an uncertain outcome
 - ⇒ The probability of an event **measures** how likely it is to occur

Example

- ▶ I've written a student's name in a piece of paper. Who is she/he?
- ▶ **Event:** Student x 's name is written in the paper
- ▶ **Probability:** $P(x)$ measures how likely it is that x 's name was written
- ▶ **Probability is a measurement tool**
 - ⇒ Mathematical language for quantifying uncertainty

- ▶ Given a **sample space** or universe S
 - ▶ **Ex:** All students in the class $S = \{x_1, x_2, \dots, x_N\}$ (x_n denote names)
- ▶ **Def:** An **outcome** is an element or point in S , e.g., x_3
- ▶ **Def:** An **event** E is a subset of S
 - ▶ **Ex:** $\{x_1\}$, student with name x_1
 - ▶ **Ex:** Also $\{x_1, x_4\}$, students with names x_1 and x_4
 - ⇒ Outcome x_3 and event $\{x_3\}$ **are different**, the latter is a set
- ▶ **Def:** A **sigma-algebra** \mathcal{F} is a collection of events $E \subseteq S$ such that
 - (i) The empty set \emptyset belongs to \mathcal{F} : $\emptyset \in \mathcal{F}$
 - (ii) Closed under **complement**: If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
 - (iii) Closed under **countable unions**: If $E_1, E_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$
- ▶ \mathcal{F} is a set of sets

Example

- ▶ No student and all students, i.e., $\mathcal{F}_0 := \{\emptyset, S\}$

Example

- ▶ Empty set, women, men, everyone, i.e., $\mathcal{F}_1 := \{\emptyset, \text{Women}, \text{Men}, S\}$

Example

- ▶ \mathcal{F}_2 including the empty set \emptyset **plus**

All events (sets) with one student $\{x_1\}, \dots, \{x_N\}$ **plus**

All events with two students $\{x_1, x_2\}, \{x_1, x_3\}, \dots, \{x_1, x_N\},$
 $\{x_2, x_3\}, \dots, \{x_2, x_N\},$

...

$\{x_{N-1}, x_N\}$ **plus**

All events with three, four, ..., N students

$\Rightarrow \mathcal{F}_2$ is known as the **power set** of S , denoted 2^S

- ▶ Define a function $P(E)$ from a sigma-algebra \mathcal{F} to the real numbers
- ▶ $P(E)$ qualifies as a probability if
 - A1) **Non-negativity**: $P(E) \geq 0$
 - A2) **Probability of universe**: $P(S) = 1$
 - A3) **Additivity**: Given sequence of **disjoint** events E_1, E_2, \dots

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

⇒ Disjoint (mutually exclusive) events means $E_i \cap E_j = \emptyset$, $i \neq j$

⇒ Union of countably infinite many disjoint events

- ▶ Triplet $(S, \mathcal{F}, P(\cdot))$ is called a **probability space**

► Implications of the axioms A1)-A3)

⇒ **Impossible event**: $P(\emptyset) = 0$

⇒ **Monotonicity**: $E_1 \subset E_2 \Rightarrow P(E_1) \leq P(E_2)$

⇒ **Range**: $0 \leq P(E) \leq 1$

⇒ **Complement**: $P(E^c) = 1 - P(E)$

⇒ **Finite disjoint union**: For disjoint events E_1, \dots, E_N

$$P\left(\bigcup_{i=1}^N E_i\right) = \sum_{i=1}^N P(E_i)$$

⇒ **Inclusion-exclusion**: For **any** events E_1 and E_2

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

- ▶ Let's construct a **probability space** for our running example
- ▶ Universe of all students in the class $S = \{x_1, x_2, \dots, x_N\}$
- ▶ Sigma-algebra with all combinations of students, i.e., $\mathcal{F} = 2^S$
- ▶ Suppose names are equiprobable $\Rightarrow P(\{x_n\}) = 1/N$ for all n
 \Rightarrow Have to **specify probability for all $E \in \mathcal{F}$** \Rightarrow Define $P(E) = \frac{|E|}{|S|}$
- ▶ **Q:** Is this function a probability?
 \Rightarrow **A1):** $P(E) = \frac{|E|}{|S|} \geq 0 \checkmark \Rightarrow$ **A2):** $P(S) = \frac{|S|}{|S|} = 1 \checkmark$
 \Rightarrow **A3):** $P\left(\bigcup_{i=1}^N E_i\right) = \frac{|\bigcup_{i=1}^N E_i|}{|S|} = \frac{\sum_{i=1}^N |E_i|}{|S|} = \sum_{i=1}^N P(E_i) \checkmark$
- ▶ The $P(\cdot)$ just defined is called **uniform probability distribution**

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

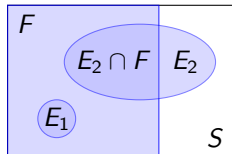
Joint expectations

- ▶ Consider events E and F , and **suppose we know F occurred**
- ▶ **Q:** What does this information imply about the probability of E ?
- ▶ **Def:** **Conditional probability of E given F** is (need $P(F) > 0$)

$$P(E | F) = \frac{P(E \cap F)}{P(F)}$$

\Rightarrow In general $P(E|F) \neq P(F|E)$

- ▶ **Renormalize** probabilities to the set F
 - ▶ Discard a piece of S
 - ▶ May discard a piece of E as well



- ▶ For given F with $P(F) > 0$, $P(\cdot|F)$ satisfies the axioms of probability

Conditional probability example

- ▶ The name I wrote is male. What is the probability of name x_n ?
- ▶ Assume male names are $F = \{x_1, \dots, x_M\} \Rightarrow P(F) = \frac{M}{N}$
- ▶ If name x_n is **male**, $x_n \in F$ and we have for event $E = \{x_n\}$

$$P(E \cap F) = P(\{x_n\}) = \frac{1}{N}$$

\Rightarrow Conditional probability is as you would expect

$$P(E | F) = \frac{P(E \cap F)}{P(F)} = \frac{1/N}{M/N} = \frac{1}{M}$$

- ▶ If name is **female** $x_n \notin F$, then $P(E \cap F) = P(\emptyset) = 0$
 \Rightarrow As you would expect, then $P(E | F) = 0$

- ▶ Consider event E and events F and F^c
 - ▶ F and F^c form a **partition** of the space S ($F \cup F^c = S$, $F \cap F^c = \emptyset$)

- ▶ Because $F \cup F^c = S$ cover space S , can write the set E as

$$E = E \cap S = E \cap [F \cup F^c] = [E \cap F] \cup [E \cap F^c]$$

- ▶ Because $F \cap F^c = \emptyset$ are **disjoint**, so is $[E \cap F] \cap [E \cap F^c] = \emptyset$
 $\Rightarrow P(E) = P([E \cap F] \cup [E \cap F^c]) = P(E \cap F) + P(E \cap F^c)$

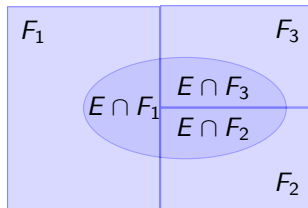
- ▶ Use definition of conditional probability

$$P(E) = P(E \mid F)P(F) + P(E \mid F^c)P(F^c)$$

- ▶ Translate **conditional** information $P(E \mid F)$ and $P(E \mid F^c)$
 \Rightarrow Into **unconditional** information $P(E)$

Law of total probability (continued)

- ▶ In general, consider (possibly infinite) **partition** F_i , $i = 1, 2, \dots$ of S
- ▶ Sets are **disjoint** $\Rightarrow F_i \cap F_j = \emptyset$ for $i \neq j$
- ▶ Sets **cover the space** $\Rightarrow \bigcup_{i=1}^{\infty} F_i = S$



- ▶ As before, because $\bigcup_{i=1}^{\infty} F_i = S$ **cover the space**, can write set E as

$$E = E \cap S = E \cap \left[\bigcup_{i=1}^{\infty} F_i \right] = \bigcup_{i=1}^{\infty} [E \cap F_i]$$

- ▶ Because $F_i \cap F_j = \emptyset$ are **disjoint**, so is $[E \cap F_i] \cap [E \cap F_j] = \emptyset$. Thus

$$P(E) = P\left(\bigcup_{i=1}^{\infty} [E \cap F_i]\right) = \sum_{i=1}^{\infty} P(E \cap F_i) = \sum_{i=1}^{\infty} P(E | F_i) P(F_i)$$

- ▶ Consider a probability class in some university
 - ⇒ Seniors get an A with probability (w.p.) 0.9, juniors w.p. 0.8
 - ⇒ An exchange student is a senior w.p. 0.7, and a junior w.p. 0.3
- ▶ Q: What is the probability of the exchange student scoring an A?
- ▶ Let A = “exchange student gets an A,” S denote senior, and J junior
 - ⇒ Use the **law of total probability**

$$\begin{aligned}P(A) &= P(A \mid S)P(S) + P(A \mid J)P(J) \\ &= 0.9 \times 0.7 + 0.8 \times 0.3 = 0.87\end{aligned}$$

- ▶ From the definition of conditional probability

$$P(E | F)P(F) = P(E \cap F)$$

- ▶ Likewise, for F conditioned on E we have

$$P(F | E)P(E) = P(F \cap E)$$

- ▶ Quantities above are equal, giving **Bayes' rule**

$$P(E | F) = \frac{P(F | E)P(E)}{P(F)}$$

- ▶ Bayes' rule allows **time reversion**. If F (future) comes after E (past),
 - $\Rightarrow P(E | F)$, probability of past (E) having seen the future (F)
 - $\Rightarrow P(F | E)$, probability of future (F) having seen past (E)
- ▶ Models often describe **future | past**. Interest is often in **past | future**

- ▶ Consider the following partition of my email
 - $\Rightarrow E_1 = \text{"spam"} \text{ w.p. } P(E_1) = 0.7$
 - $\Rightarrow E_2 = \text{"low priority"} \text{ w.p. } P(E_2) = 0.2$
 - $\Rightarrow E_3 = \text{"high priority"} \text{ w.p. } P(E_3) = 0.1$
- ▶ Let $F = \text{"an email contains the word free"}$
 - \Rightarrow From experience know $P(F | E_1) = 0.9, P(F | E_2) = P(F | E_3) = 0.01$
- ▶ I got an email containing "free". What is the probability that it is spam?
- ▶ Apply **Bayes' rule**

$$P(E_1 | F) = \frac{P(F | E_1)P(E_1)}{P(F)} = \frac{P(F | E_1)P(E_1)}{\sum_{i=1}^3 P(F | E_i)P(E_i)} = 0.995$$

\Rightarrow **Law of total probability** very useful when applying Bayes' rule

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ **Def:** Events E and F are **independent** if $P(E \cap F) = P(E)P(F)$

⇒ Events that are not independent are **dependent**

- ▶ According to definition of conditional probability

$$P(E | F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E)$$

⇒ Intuitive, **knowing F does not alter our perception of E**

⇒ **F bears no information about E**

⇒ The symmetric is also true $P(F | E) = P(F)$

- ▶ Whether E and F are independent relies strongly on $P(\cdot)$
- ▶ Avoid confusing with disjoint events, meaning $E \cap F = \emptyset$
- ▶ **Q:** Can disjoint events with $P(E) > 0$, $P(F) > 0$ be independent? **No**

- ▶ Wrote one name, asked a friend to write another (possibly the same)
- ▶ Probability space $(S, \mathcal{F}, P(\cdot))$ for this experiment
 - ⇒ S is the set of all pairs of names $[x_n(1), x_n(2)]$, $|S| = N^2$
 - ⇒ Sigma-algebra is cartesian product $\mathcal{F} = 2^S \times 2^S$
 - ⇒ Define $P(E) = \frac{|E|}{|S|}$ as the uniform probability distribution
- ▶ Consider the events $E_1 = \text{'I wrote } x_1 \text{'}$ and $E_2 = \text{'My friend wrote } x_2 \text{'}$
Q: Are they **independent**? **Yes**, since

$$P(E_1 \cap E_2) = P(\{(x_1, x_2)\}) = \frac{|\{(x_1, x_2)\}|}{|S|} = \frac{1}{N^2} = P(E_1)P(E_2)$$

- ▶ **Dependent** events: $E_1 = \text{'I wrote } x_1 \text{'}$ and $E_3 = \text{'Both names are male'}$

- **Def:** Events E_i , $i = 1, 2, \dots$ are called **mutually independent** if

$$P\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} P(E_i)$$

for **every finite** subset I of at least two integers

- **Ex:** Events E_1 , E_2 , and E_3 are mutually independent if all the following hold

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$$

$$P(E_1 \cap E_2) = P(E_1)P(E_2)$$

$$P(E_1 \cap E_3) = P(E_1)P(E_3)$$

$$P(E_2 \cap E_3) = P(E_2)P(E_3)$$

- If $P(E_i \cap E_j) = P(E_i)P(E_j)$ for all (i, j) , the E_i are **pairwise independent**
⇒ Mutual independence → pairwise independence. **Not the other way**

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ **Def:** RV $X(s)$ is a **function** that assigns a value to an outcome $s \in S$
⇒ Think of RVs as measurements associated with an experiment

Example

- ▶ Throw a ball inside a $1m \times 1m$ square. Interested in ball position
- ▶ **Uncertain outcome** is the place s where the ball falls
- ▶ **Random variables** are $X(s)$ and $Y(s)$ position coordinates

- ▶ RV probabilities inferred from probabilities of underlying outcomes

$$P(X(s) = x) = P(\{s \in S : X(s) = x\})$$

$$P(X(s) \in (-\infty, x]) = P(\{s \in S : X(s) \in (-\infty, x]\})$$

Example 1

- ▶ Throw coin for head (H) or tails (T). Coin is fair $P(H) = 1/2$, $P(T) = 1/2$. Pay \$1 for H , charge \$1 for T . Earnings?
- ▶ Possible outcomes are H and T
- ▶ To measure earnings define RV X with values

$$X(H) = 1, \quad X(T) = -1$$

- ▶ Probabilities of the RV are

$$P(X = 1) = P(H) = 1/2,$$

$$P(X = -1) = P(T) = 1/2$$

⇒ Also have $P(X = x) = 0$ for all other $x \neq \pm 1$

Example 2

- ▶ Throw 2 coins. Pay \$1 for each H , charge \$1 for each T . Earnings?
- ▶ Now the possible outcomes are HH , HT , TH , and TT
- ▶ To measure earnings define RV Y with values

$$Y(HH) = 2, \quad Y(HT) = 0, \quad Y(TH) = 0, \quad Y(TT) = -2$$

- ▶ Probabilities of the RV are

$$P(Y = 2) = P(HH) = 1/4,$$

$$P(Y = 0) = P(HT) + P(TH) = 1/2,$$

$$P(Y = -2) = P(TT) = 1/4$$

- ▶ RVs are easier to manipulate than events
- ▶ Let $s_1 \in \{H, T\}$ be outcome of coin 1 and $s_2 \in \{H, T\}$ of coin 2
 \Rightarrow Can relate Y and X s as

$$Y(s_1, s_2) = X_1(s_1) + X_2(s_2)$$

- ▶ Throw N coins. **Earnings?** Enumeration becomes cumbersome
- ▶ Alternatively, let $s_n \in \{H, T\}$ be outcome of n -th toss and define

$$Y(s_1, s_2, \dots, s_N) = \sum_{n=1}^N X_n(s_n)$$

\Rightarrow Will usually abuse notation and write $Y = \sum_{n=1}^N X_n$

Example 3

- ▶ Throw a coin until landing heads for the first time. $P(H) = p$
- ▶ Number of throws until the first head?
- ▶ Outcomes are $H, TH, TTH, TTTH, \dots$. Note that $|S| = \infty$
 \Rightarrow Stop tossing after first H (thus THT not a possible outcome)
- ▶ Let N be a RV counting the number of throws
 $\Rightarrow N = n$ if we land T in the first $n - 1$ throws and H in the n -th

$$\begin{aligned}P(N = 1) &= P(H) &&= p \\P(N = 2) &= P(TH) &&= (1 - p)p \\&\vdots \\P(N = n) &= P(\underbrace{TT \dots T}_{n-1 \text{ tails}} H) &&= (1 - p)^{n-1}p\end{aligned}$$

Example 3 (continued)

- ▶ From **A2)** we should have $P(S) = \sum_{n=1}^{\infty} P(N=n) = 1$
- ▶ Holds because $\sum_{n=1}^{\infty} (1-p)^{n-1}$ is a **geometric series**

$$\sum_{n=1}^{\infty} (1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \dots = \frac{1}{1-(1-p)} = \frac{1}{p}$$

- ▶ Plug the sum of the geometric series in the expression for $P(S)$

$$\sum_{n=1}^{\infty} P(N=n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = p \times \frac{1}{p} = 1 \checkmark$$

- ▶ The **indicator function of an event** is a random variable
- ▶ Let $s \in S$ be an outcome, and $E \subset S$ be an event

$$\mathbb{I}\{E\}(s) = \begin{cases} 1, & \text{if } s \in E \\ 0, & \text{if } s \notin E \end{cases}$$

⇒ Indicates that outcome s belongs to set E , by taking value 1

Example

- ▶ Number of throws N until first H. Interested on N exceeding N_0
 - ⇒ Event is $\{N : N > N_0\}$. Possible outcomes are $N = 1, 2, \dots$
 - ⇒ Denote indicator function as $\mathbb{I}_{N_0} = \mathbb{I}\{N : N > N_0\}$
- ▶ Probability $P(\mathbb{I}_{N_0} = 1) = P(N > N_0) = (1 - p)^{N_0}$
 - ⇒ For N to exceed N_0 need N_0 consecutive tails
 - ⇒ **Doesn't matter what happens afterwards**

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

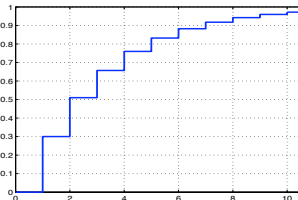
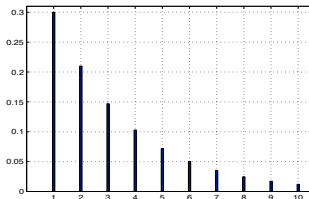
Joint probability distributions

Joint expectations

- ▶ **Discrete RV** takes on, at most, a **countable** number of values
- ▶ **Probability mass function (pmf)** $p_X(x) = P(X = x)$
 - ▶ If RV is clear from context, just write $p_X(x) = p(x)$
- ▶ If X supported in $\{x_1, x_2, \dots\}$, pmf satisfies
 - $p(x_i) > 0$ for $i = 1, 2, \dots$
 - $p(x) = 0$ for all other $x \neq x_i$
 - $\sum_{i=1}^{\infty} p(x_i) = 1$
 - ▶ Pmf for “throw to first heads” ($p = 0.3$)
- ▶ **Cumulative distribution function (cdf)**
$$F_X(x) = P(X \leq x) = \sum_{i: x_i \leq x} p(x_i)$$

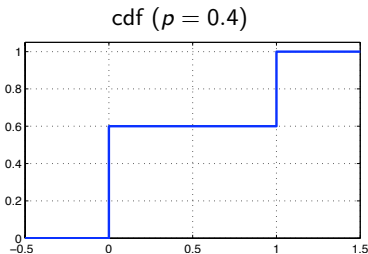
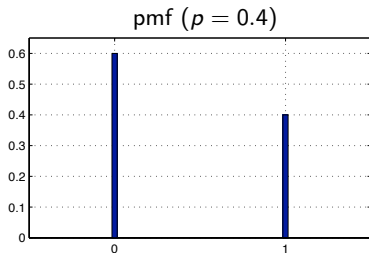
\Rightarrow **Staircase function with jumps at x_i**

 - ▶ Cdf for “throw to first heads” ($p = 0.3$)



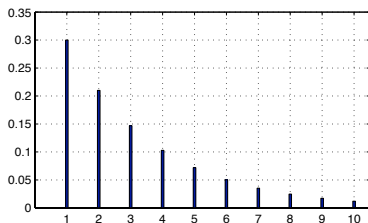
- ▶ A trial/experiment/bet can succeed w.p. p or fail w.p. $q := 1 - p$
⇒ Ex: coin throws, any indication of an event
- ▶ Bernoulli X can be 0 or 1. Pmf is $p(x) = p^x q^{1-x}$
- ▶ Cdf is

$$F(x) = \begin{cases} 0, & x < 0 \\ q, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

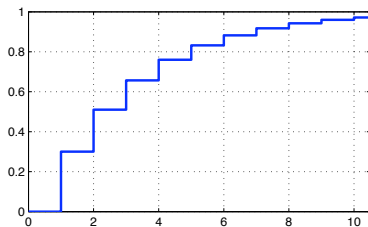


- ▶ Count number of Bernoulli trials needed to register first success
⇒ Trials succeed w.p. p
- ▶ Number of trials X until success is **geometric** with parameter p
- ▶ Pmf is $p(x) = p(1 - p)^{x-1}$
 - ▶ One success after $x - 1$ failures, trials are independent
- ▶ Cdf is $F(x) = 1 - (1 - p)^x$
 - ▶ Recall $P(X > x) = (1 - p)^x$; or just sum the geometric series

pmf ($p = 0.3$)



cdf ($p = 0.3$)



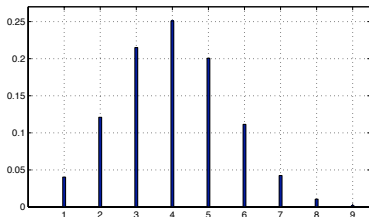
- ▶ Count number of successes X in n Bernoulli trials
 \Rightarrow Trials succeed w.p. p
- ▶ Number of successes X is binomial with parameters (n, p) . Pmf is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

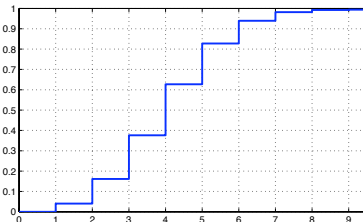
$\Rightarrow X = x$ for x successes (p^x) and $n - x$ failures ($(1-p)^{n-x}$).

$\Rightarrow \binom{n}{x}$ ways of drawing x successes and $n - x$ failures

pmf ($n = 9, p = 0.4$)



cdf ($n = 9, p = 0.4$)



- ▶ Let $Y_i, i = 1, \dots, n$ be Bernoulli RVs with parameter p
 $\Rightarrow Y_i$ associated with independent events
- ▶ Can write binomial X with parameters (n, p) as $\Rightarrow X = \sum_{i=1}^n Y_i$

Example

- ▶ Consider binomials Y and Z with parameters (n_Y, p) and (n_Z, p)
 \Rightarrow Q: Probability distribution of $X = Y + Z$?
- ▶ Write $Y = \sum_{i=1}^{n_Y} Y_i$ and $Z = \sum_{i=1}^{n_Z} Z_i$, thus

$$X = \sum_{i=1}^{n_Y} Y_i + \sum_{i=1}^{n_Z} Z_i$$

$\Rightarrow X$ is binomial with parameter $(n_Y + n_Z, p)$

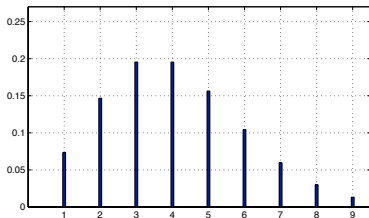
- ▶ Counts of rare events (radioactive decay, packet arrivals, accidents)
- ▶ Usually modeled as Poisson with parameter λ and pmf

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

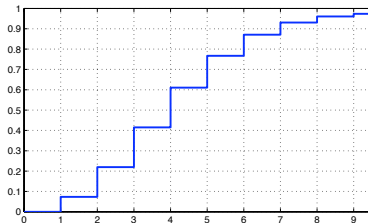
- ▶ Q: Is this a properly defined pmf? Yes
- ▶ Taylor's expansion of $e^x = 1 + x + x^2/2 + \dots + x^i/i! + \dots$. Then

$$P(S) = \sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1 \checkmark$$

pmf ($\lambda = 4$)

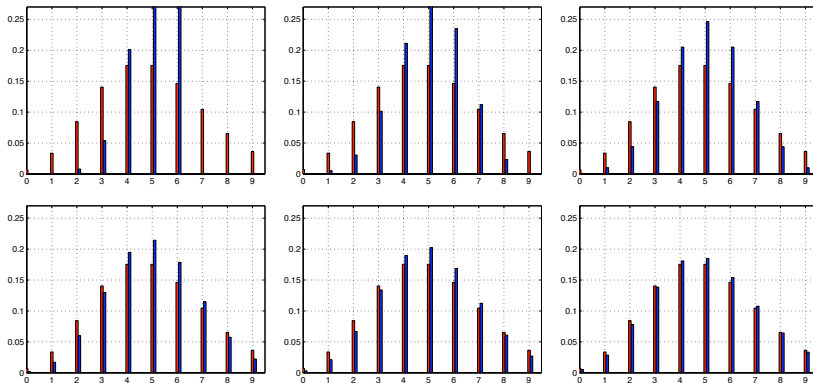


cdf ($\lambda = 4$)



Poisson approximation of binomial

- ▶ X is binomial with parameters (n, p)
- ▶ Let $n \rightarrow \infty$ while maintaining a constant product $np = \lambda$
 - ▶ If we just let $n \rightarrow \infty$ number of successes diverges. Boring
- ▶ Compare with Poisson distribution with parameter λ
 - ▶ $\lambda = 5$, $n = 6, 8, 10, 15, 20, 50$



- ▶ This is, in fact, the motivation for the definition of a Poisson RV
- ▶ Substituting $p = \lambda/n$ in the pmf of a binomial RV

$$\begin{aligned} p_n(x) &= \frac{n!}{(n-x)!x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{n(n-1)\dots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^x} \end{aligned}$$

\Rightarrow Used factorials' defs., $(1 - \lambda/n)^{n-x} = \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^x}$, and reordered terms

- ▶ In the limit, **red** term is $\lim_{n \rightarrow \infty} (1 - \lambda/n)^n = e^{-\lambda}$
- ▶ Black and **blue** terms converge to 1. From both observations

$$\lim_{n \rightarrow \infty} p_n(x) = 1 \frac{\lambda^x}{x!} \frac{e^{-\lambda}}{1} = e^{-\lambda} \frac{\lambda^x}{x!}$$

\Rightarrow **Limit is the pmf of a Poisson RV**

- ▶ Binomial distribution is motivated by counting successes
- ▶ The Poisson is an approximation for large number of trials n
 - ⇒ Poisson distribution is more tractable (compare pmfs)
- ▶ Sometimes called “law of rare events”
 - ▶ Individual events (successes) happen with small probability $p = \lambda/n$
 - ▶ Aggregate event (number of successes), though, need not be rare
- ▶ Notice that all four RVs seen so far are related to “coin tosses”

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ Possible values for continuous RV X form a dense subset $\mathcal{X} \subseteq \mathbb{R}$
 \Rightarrow **Uncountably** infinite number of possible values

- ▶ Probability density function (pdf) $f_X(x)$ is such that for any subset $\mathcal{X} \subseteq \mathbb{R}$
(Normal pdf to the right)

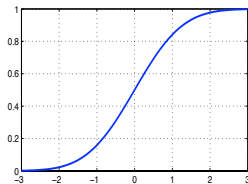
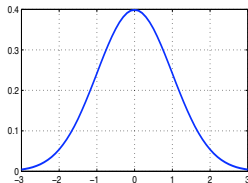
$$P(X \in \mathcal{X}) = \int_{\mathcal{X}} f_X(x) dx$$

\Rightarrow **Will have** $P(X = x) = 0$ for all $x \in \mathcal{X}$

- ▶ Cdf defined as before and related to the pdf
(Normal cdf to the right)

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du$$

$$\Rightarrow P(X \leq \infty) = F_X(\infty) = \lim_{x \rightarrow \infty} F_X(x) = 1$$



- ▶ When the set $\mathcal{X} = [a, b]$ is an interval of \mathbb{R}

$$P(X \in [a, b]) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$$

- ▶ In terms of the pdf it can be written as

$$P(X \in [a, b]) = \int_a^b f_X(x) dx$$

- ▶ For small interval $[x_0, x_0 + \delta x]$, in particular

$$P(X \in [x_0, x_0 + \delta x]) = \int_{x_0}^{x_0 + \delta x} f_X(x) dx \approx f_X(x_0) \delta x$$

\Rightarrow Probability is the “area under the pdf” (thus “density”)

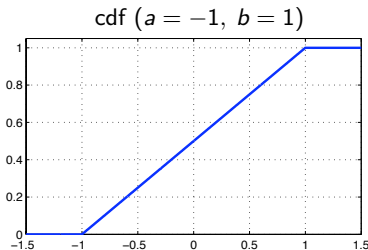
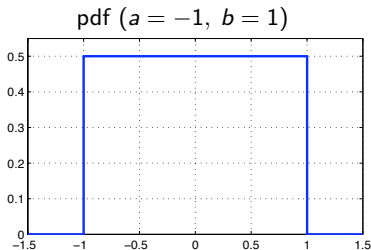
- ▶ Another relationship between pdf and cdf is $\Rightarrow \frac{\partial F_X(x)}{\partial x} = f_X(x)$

\Rightarrow Fundamental theorem of calculus (“derivative inverse of integral”)

- ▶ Model problems with equal probability of landing on an interval $[a, b]$
- ▶ Pdf of **uniform** RV is $f(x) = 0$ outside the interval $[a, b]$ and

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \leq x \leq b$$

- ▶ Cdf is $F(x) = (x-a)/(b-a)$ in the interval $[a, b]$ (0 before, 1 after)
- ▶ Prob. of interval $[\alpha, \beta] \subseteq [a, b]$ is $\int_{\alpha}^{\beta} f(x)dx = (\beta - \alpha)/(b - a)$
⇒ Depends on interval's width $\beta - \alpha$ only, not on its position



- Model duration of phone calls, lifetime of electronic components
- Pdf of **exponential** RV is

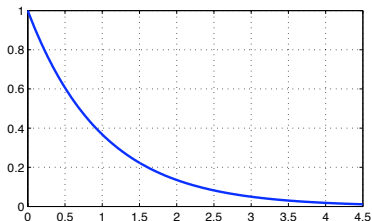
$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

⇒ As parameter λ increases, “height” increases and “width” decreases

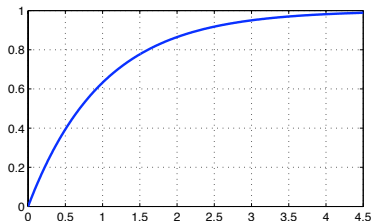
- Cdf obtained by integrating pdf

$$F(x) = \int_{-\infty}^x f(u) du = \int_0^x \lambda e^{-\lambda u} du = -e^{-\lambda u} \Big|_0^x = 1 - e^{-\lambda x}$$

pdf ($\lambda = 1$)



cdf ($\lambda = 1$)

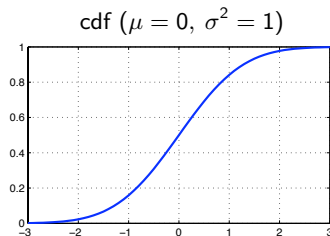
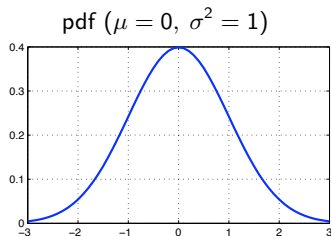


- ▶ Model randomness arising from large number of random effects
- ▶ Pdf of normal RV is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

- $\Rightarrow \mu$ is the mean (center), σ^2 is the variance (width)
- \Rightarrow 0.68 prob. between $\mu \pm \sigma$, 0.997 prob. in $\mu \pm 3\sigma$
- \Rightarrow **Standard normal** RV has $\mu = 0$ and $\sigma^2 = 1$

- ▶ Cdf $F(x)$ cannot be expressed in terms of elementary functions



Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ We are asked to summarize information about a RV in a single value
 - ⇒ What should this value be?
- ▶ If we are allowed a description with a few values
 - ⇒ What should they be?
- ▶ Expected (mean) values are convenient answers to these questions
- ▶ **Beware:** Expectations are condensed descriptions
 - ⇒ They overlook some aspects of the random phenomenon
 - ⇒ Whole story told by the probability distribution (cdf)

- ▶ Discrete RV X taking on values x_i , $i = 1, 2, \dots$ with pmf $p(x)$
- ▶ **Def:** The **expected value** of the **discrete** RV X is

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p(x_i) = \sum_{x:p(x)>0} xp(x)$$

- ▶ Weighted average of possible values x_i . **Probabilities are weights**
- ▶ Common average if RV takes values x_i , $i = 1, \dots, N$ equiprobably

$$\mathbb{E}[X] = \sum_{i=1}^N x_i p(x_i) = \sum_{i=1}^N x_i \frac{1}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

Ex: For a **Bernoulli** RV $p(x) = p^x q^{1-x}$, for $x \in \{0, 1\}$

$$\mathbb{E}[X] = 1 \times p + 0 \times q = p$$

Ex: For a **geometric** RV $p(x) = p(1-p)^{x-1} = pq^{x-1}$, for $x \geq 1$

- Note that $\partial q^x / \partial q = xq^{x-1}$ and that derivatives are linear operators

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x p q^{x-1} = p \sum_{x=1}^{\infty} \frac{\partial q^x}{\partial q} = p \frac{\partial}{\partial q} \left(\sum_{x=1}^{\infty} q^x \right)$$

- Sum inside derivative is geometric. Sums to $q/(1-q)$, thus

$$\mathbb{E}[X] = p \frac{\partial}{\partial q} \left(\frac{q}{1-q} \right) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

- Time to first success is inverse of success probability. Reasonable

Ex: For a **Poisson** RV $p(x) = e^{-\lambda}(\lambda^x/x!)$, for $x \geq 0$

- ▶ First summand in definition is 0, pull λ out, and use $\frac{x}{x!} = \frac{1}{(x-1)!}$

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

- ▶ Sum is Taylor's expansion of $e^{\lambda} = 1 + \lambda + \lambda^2/2! + \dots + \lambda^x/x!$

$$\mathbb{E}[X] = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

- ▶ Poisson is limit of binomial for large number of trials n , with $\lambda = np$
 - \Rightarrow Counts number of successes in n trials that succeed w.p. p
- ▶ Expected number of successes is $\lambda = np$
 - \Rightarrow Number of trials \times probability of individual success. Reasonable

- ▶ Continuous RV X taking values on \mathbb{R} with pdf $f(x)$
- ▶ **Def:** The **expected value** of the **continuous** RV X is

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} xf(x) dx$$

- ▶ Compare with $\mathbb{E}[X] := \sum_{x:p(x)>0} xp(x)$ in the discrete RV case
- ▶ Note that the integral or sum are assumed to be well defined
⇒ Otherwise we say the **expectation does not exist**

Ex: For a **normal** RV add and subtract μ , separate integrals

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x + \mu - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \mu \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx\end{aligned}$$

- ▶ **First integral** is 1 because it integrates a pdf in all \mathbb{R}
- ▶ **Second integral** is 0 by symmetry. Both observations yield

$$\mathbb{E}[X] = \mu$$

- ▶ The mean of a RV with a symmetric pdf is the point of symmetry

Ex: For a **uniform** RV $f(x) = 1/(b - a)$, for $a \leq x \leq b$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{(a+b)}{2}$$

- Makes sense, since pdf is **symmetric** around midpoint $(a+b)/2$

Ex: For an **exponential** RV (non symmetric) integrate by parts

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{\infty} x\lambda e^{-\lambda x} dx \\ &= -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= -xe^{-\lambda x} \Big|_0^{\infty} - \frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = \frac{1}{\lambda}\end{aligned}$$

- ▶ Consider a function $g(X)$ of a RV X . Expected value of $g(X)$?
- ▶ $g(X)$ is also a RV, then it also has a pmf $p_{g(X)}(g(x))$

$$\mathbb{E}[g(X)] = \sum_{g(x): p_{g(X)}(g(x)) > 0} g(x) p_{g(X)}(g(x))$$

⇒ Requires calculating the pmf of $g(X)$. There is a simpler way

Theorem

Consider a function $g(X)$ of a discrete RV X with pmf $p_X(x)$. Then

$$\mathbb{E}[g(X)] = \sum_{i=1}^{\infty} g(x_i) p_X(x_i)$$

- ▶ Weighted average of functional values. No need to find pmf of $g(X)$
- ▶ Same can be proved for a continuous RV

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

- Consider a **linear function** (actually affine) $g(X) = aX + b$

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_{i=1}^{\infty} (ax_i + b)p_X(x_i) \\ &= \sum_{i=1}^{\infty} ax_i p_X(x_i) + \sum_{i=1}^{\infty} bp_X(x_i) \\ &= a \sum_{i=1}^{\infty} x_i p_X(x_i) + b \sum_{i=1}^{\infty} p_X(x_i) \\ &= a\mathbb{E}[X] + b1\end{aligned}$$

- Can interchange expectation with additive/multiplicative constants

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

⇒ Again, the same holds for a continuous RV

- ▶ Let X be a RV and \mathcal{X} be a set

$$\mathbb{I}\{X \in \mathcal{X}\} = \begin{cases} 1, & \text{if } x \in \mathcal{X} \\ 0, & \text{if } x \notin \mathcal{X} \end{cases}$$

- ▶ Expected value of $\mathbb{I}\{X \in \mathcal{X}\}$ in the discrete case

$$\mathbb{E}[\mathbb{I}\{X \in \mathcal{X}\}] = \sum_{x: p_X(x) > 0} \mathbb{I}\{x \in \mathcal{X}\} p_X(x) = \sum_{x \in \mathcal{X}} p_X(x) = \mathbf{P}(X \in \mathcal{X})$$

- ▶ Likewise in the continuous case

$$\mathbb{E}[\mathbb{I}\{X \in \mathcal{X}\}] = \int_{-\infty}^{\infty} \mathbb{I}\{x \in \mathcal{X}\} f_X(x) dx = \int_{x \in \mathcal{X}} f_X(x) dx = \mathbf{P}(X \in \mathcal{X})$$

- ▶ Expected value of indicator RV = Probability of indicated event
⇒ Recall $\mathbb{E}[X] = p$ for Bernoulli RV (it “indicates success”)

- **Def:** The n -th moment ($n \geq 0$) of a RV is

$$\mathbb{E}[X^n] = \sum_{i=1}^{\infty} x_i^n p(x_i)$$

- **Def:** The n -th central moment corrects for the mean, that is

$$\mathbb{E}\left[(X - \mathbb{E}[X])^n\right] = \sum_{i=1}^{\infty} (x_i - \mathbb{E}[X])^n p(x_i)$$

- 0-th order moment is $\mathbb{E}[X^0] = 1$; 1-st moment is the mean $\mathbb{E}[X]$
- 2-nd central moment is the **variance**. Measures **width of the pmf**

$$\text{var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Ex: For affine functions

$$\text{var}[aX + b] = a^2 \text{var}[X]$$

Ex: For a **Bernoulli** RV X with parameter p , $\mathbb{E}[X] = \mathbb{E}[X^2] = p$
 $\Rightarrow \text{var}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = p - p^2 = p(1 - p)$

Ex: For **Poisson** RV Y with parameter λ , second moment is

$$\begin{aligned}\mathbb{E}[Y^2] &= \sum_{y=0}^{\infty} y^2 e^{-\lambda} \frac{\lambda^y}{y!} = \sum_{y=1}^{\infty} y \frac{e^{-\lambda} \lambda^y}{(y-1)!} \\&= \sum_{y=1}^{\infty} (y-1) \frac{e^{-\lambda} \lambda^y}{(y-1)!} + \sum_{y=1}^{\infty} \frac{e^{-\lambda} \lambda^y}{(y-1)!} \\&= e^{-\lambda} \lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} + e^{-\lambda} \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} \\&= e^{-\lambda} \lambda^2 e^{\lambda} + e^{-\lambda} \lambda e^{\lambda} = \lambda^2 + \lambda\end{aligned}$$

$$\Rightarrow \text{var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}^2[Y] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ Want to study problems with more than one RV. Say, e.g., X and Y
- ▶ Probability distributions of X and Y **are not sufficient**
⇒ **Joint probability distribution (cdf) of (X, Y)** defined as

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

- ▶ If X, Y clear from context omit subindex to write $F_{XY}(x, y) = F(x, y)$
- ▶ Can recover $F_X(x)$ by considering all possible values of Y

$$F_X(x) = P(X \leq x) = P(X \leq x, Y \leq \infty) = F_{XY}(x, \infty)$$

⇒ $F_X(x)$ and $F_Y(y) = F_{XY}(\infty, y)$ are called **marginal cdfs**

- ▶ Consider discrete RVs X and Y
 X takes values in $\mathcal{X} := \{x_1, x_2, \dots\}$ and Y in $\mathcal{Y} := \{y_1, y_2, \dots\}$

- ▶ **Joint pmf** of (X, Y) defined as

$$p_{XY}(x, y) = P(X = x, Y = y)$$

- ▶ Possible values (x, y) are elements of the Cartesian product $\mathcal{X} \times \mathcal{Y}$
 - ▶ $(x_1, y_1), (x_1, y_2), \dots, (x_2, y_1), (x_2, y_2), \dots, (x_3, y_1), (x_3, y_2), \dots$
- ▶ Marginal pmf $p_X(x)$ obtained by summing over all values of Y

$$p_X(x) = P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x, Y = y) = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$$

\Rightarrow Likewise $p_Y(y) = \sum_{x \in \mathcal{X}} p_{XY}(x, y)$. **Marginalize by summing**

- ▶ Consider continuous RVs X, Y . Arbitrary set $\mathcal{A} \in \mathbb{R}^2$
- ▶ **Joint pdf** is a function $f_{XY}(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ such that

$$P((X, Y) \in \mathcal{A}) = \iint_{\mathcal{A}} f_{XY}(x, y) dx dy$$

- ▶ **Marginalization**. There are two ways of writing $P(X \in \mathcal{X})$

$$P(X \in \mathcal{X}) = P(X \in \mathcal{X}, Y \in \mathbb{R}) = \int_{\mathcal{X}} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx$$

$$\Rightarrow \text{Definition of } f_X(x) \Rightarrow P(X \in \mathcal{X}) = \int_{\mathcal{X}} f_X(x) dx$$

- ▶ Lipstick on a pig (same thing written differently is still same thing)

$$\Rightarrow f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

- ▶ Consider two Bernoulli RVs B_1, B_2 , with the same parameter p
 \Rightarrow Define $X = B_1$ and $Y = B_1 + B_2$

- ▶ The pmf of X is

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

- ▶ Likewise, the pmf of Y is

$$p_Y(0) = (1 - p)^2, \quad p_Y(1) = 2p(1 - p), \quad p_Y(2) = p^2$$

- ▶ The joint pmf of X and Y is

$$\begin{aligned} p_{XY}(0,0) &= (1 - p)^2, & p_{XY}(0,1) &= p(1 - p), & p_{XY}(0,2) &= 0 \\ p_{XY}(1,0) &= 0, & p_{XY}(1,1) &= p(1 - p), & p_{XY}(1,2) &= p^2 \end{aligned}$$

- ▶ For convenience often arrange RVs in a vector
⇒ Prob. distribution of vector is joint distribution of its entries

- ▶ Consider, e.g., two RVs X and Y . Random vector is $\mathbf{X} = [X, Y]^T$

- ▶ If X and Y are discrete, vector variable \mathbf{X} is discrete with pmf

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}([x, y]^T) = p_{XY}(x, y)$$

- ▶ If X, Y continuous, \mathbf{X} continuous with pdf

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}([x, y]^T) = f_{XY}(x, y)$$

- ▶ Vector cdf is ⇒ $F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}([x, y]^T) = F_{XY}(x, y)$

- ▶ In general, can define n -dimensional RVs $\mathbf{X} := [X_1, X_2, \dots, X_n]^T$
⇒ Just notation, definitions carry over from the $n = 2$ case

Sigma-algebras and probability spaces

Conditional probability, total probability, Bayes' rule

Independence

Random variables

Discrete random variables

Continuous random variables

Expected values

Joint probability distributions

Joint expectations

- ▶ RVs X and Y and function $g(X, Y)$. Function $g(X, Y)$ also a RV
- ▶ Expected value of $g(X, Y)$ when X and Y discrete can be written as

$$\mathbb{E}[g(X, Y)] = \sum_{x, y: p_{XY}(x, y) > 0} g(x, y) p_{XY}(x, y)$$

- ▶ When X and Y are continuous

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

⇒ Can have more than two RVs and use vector notation

Ex: Linear transformation of a vector RV $\mathbf{X} \in \mathbb{R}^n$: $g(\mathbf{X}) = \mathbf{a}^T \mathbf{X}$

$$\Rightarrow \mathbb{E}[\mathbf{a}^T \mathbf{X}] = \int_{\mathbb{R}^n} \mathbf{a}^T \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

- ▶ Expected value of the sum of two continuous RVs

$$\begin{aligned}\mathbb{E}[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x, y) dx dy\end{aligned}$$

- ▶ Remove x (y) from innermost integral in first (second) summand

$$\begin{aligned}\mathbb{E}[X + Y] &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

⇒ Used marginal expressions

- ▶ Expectation \leftrightarrow summation $\Rightarrow \mathbb{E}[\sum_i X_i] = \sum_i \mathbb{E}[X_i]$

- Combining with earlier result $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ proves that

$$\mathbb{E}[a_x X + a_y Y + b] = a_x \mathbb{E}[X] + a_y \mathbb{E}[Y] + b$$

- Better yet, using vector notation (with $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^n$, b a scalar)

$$\mathbb{E}[\mathbf{a}^T \mathbf{X} + b] = \mathbf{a}^T \mathbb{E}[\mathbf{X}] + b$$

- Also, if \mathbf{A} is an $m \times n$ matrix with rows $\mathbf{a}_1^T, \dots, \mathbf{a}_m^T$ and $\mathbf{b} \in \mathbb{R}^m$ a vector with elements b_1, \dots, b_m , we can write

$$\mathbb{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \begin{pmatrix} \mathbb{E}[\mathbf{a}_1^T \mathbf{X} + b_1] \\ \mathbb{E}[\mathbf{a}_2^T \mathbf{X} + b_2] \\ \vdots \\ \mathbb{E}[\mathbf{a}_m^T \mathbf{X} + b_m] \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^T \mathbb{E}[\mathbf{X}] + b_1 \\ \mathbf{a}_2^T \mathbb{E}[\mathbf{X}] + b_2 \\ \vdots \\ \mathbf{a}_m^T \mathbb{E}[\mathbf{X}] + b_m \end{pmatrix} = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}$$

- Expected value operator can be interchanged with linear operations

- ▶ Events E and F are independent if $P(E \cap F) = P(E)P(F)$
- ▶ **Def:** RVs X and Y are **independent** if events $X \leq x$ and $Y \leq y$ are independent for all x and y , i.e.

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

⇒ By definition, equivalent to $F_{XY}(x, y) = F_X(x)F_Y(y)$

- ▶ For discrete RVs equivalent to analogous relation between pmfs

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$

- ▶ For continuous RVs the analogous is true for pdfs

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

- ▶ Independence \Leftrightarrow Joint distribution factorizes into product of marginals

- ▶ **Independent** Poisson RVs X and Y with parameters λ_x and λ_y
- ▶ **Q:** Probability distribution of the sum RV $Z := X + Y$?
- ▶ $Z = n$ only if $X = k$, $Y = n - k$ for some $0 \leq k \leq n$
(use independence, Poisson pmf, rearrange terms, binomial theorem)

$$\begin{aligned} p_Z(n) &= \sum_{k=0}^n P(X = k, Y = n - k) = \sum_{k=0}^n P(X = k) P(Y = n - k) \\ &= \sum_{k=0}^n e^{-\lambda_x} \frac{\lambda_x^k}{k!} e^{-\lambda_y} \frac{\lambda_y^{n-k}}{(n-k)!} = \frac{e^{-(\lambda_x + \lambda_y)}}{n!} \sum_{k=0}^n \frac{n!}{(n-k)!k!} \lambda_x^k \lambda_y^{n-k} \\ &= \frac{e^{-(\lambda_x + \lambda_y)}}{n!} (\lambda_x + \lambda_y)^n \end{aligned}$$

- ▶ Z is Poisson with parameter $\lambda_z := \lambda_x + \lambda_y$
⇒ **Sum of independent Poissons is Poisson** (parameters added)

- ▶ Binomial RVs count number of successes in n Bernoulli trials

Ex: Let $X_i, i = 1, \dots, n$ be n **independent** Bernoulli RVs

- ▶ Can write binomial $X = \sum_{i=1}^n X_i \Rightarrow \mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np$
- ▶ **Expected nr. successes = nr. trials \times prob. individual success**
 - ▶ Same interpretation that we observed for Poisson RVs

Ex: **Dependent** Bernoulli trials. $Y = \sum_{i=1}^n X_i$, but X_i are not independent

- ▶ Expected nr. successes is still $\mathbb{E}[Y] = np$
 - ▶ **Linearity of expectation does not require independence**
 - ▶ Y is not binomial distributed

Theorem

For independent RVs X and Y , and arbitrary functions $g(X)$ and $h(Y)$:

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

The expected value of the product is the product of the expected values

- ▶ Can show that $g(X)$ and $h(Y)$ are also independent. **Intuitive**

Ex: Special case when $g(X) = X$ and $h(Y) = Y$ yields

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

- ▶ **Expectation and product can be interchanged if RVs are independent**
- ▶ Different from interchange with linear operations (**always possible**)

Proof.

- Suppose X and Y continuous RVs. Use definition of independence

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \textcolor{red}{g(x)}\textcolor{blue}{h(y)}\textcolor{red}{f_X(x)}\textcolor{blue}{f_Y(y)} dx dy\end{aligned}$$

- Integrand is product of a function of x and a function of y

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty} \textcolor{red}{g(x)}\textcolor{red}{f_X(x)} dx \int_{-\infty}^{\infty} \textcolor{blue}{h(y)}\textcolor{blue}{f_Y(y)} dy \\ &= \mathbb{E}[g(X)] \mathbb{E}[h(Y)]\end{aligned}$$



Variance of a sum of independent RVs

- ▶ Let X_n , $n = 1, \dots, N$ be independent with $\mathbb{E}[X_n] = \mu_n$, $\text{var}[X_n] = \sigma_n^2$
- ▶ **Q:** Variance of sum $X := \sum_{n=1}^N X_n$?
- ▶ Notice that mean of X is $\mathbb{E}[X] = \sum_{n=1}^N \mu_n$. Then

$$\text{var}[X] = \mathbb{E} \left[\left(\sum_{n=1}^N X_n - \sum_{n=1}^N \mu_n \right)^2 \right] = \mathbb{E} \left[\left(\sum_{n=1}^N (X_n - \mu_n) \right)^2 \right]$$

- ▶ Expand square and interchange summation and expectation

$$\text{var}[X] = \sum_{n=1}^N \sum_{m=1}^N \mathbb{E} \left[(X_n - \mu_n)(X_m - \mu_m) \right]$$

- ▶ Separate terms in sum. Then use independence and $\mathbb{E}(X_n - \mu_n) = 0$

$$\begin{aligned}\text{var}[X] &= \sum_{n=1, n \neq m}^N \sum_{m=1}^N \mathbb{E}[(X_n - \mu_n)(X_m - \mu_m)] + \sum_{n=1}^N \mathbb{E}[(X_n - \mu_n)^2] \\ &= \sum_{n=1, n \neq m}^N \sum_{m=1}^N \mathbb{E}(X_n - \mu_n)\mathbb{E}(X_m - \mu_m) + \sum_{n=1}^N \sigma_n^2 = \sum_{n=1}^N \sigma_n^2\end{aligned}$$

- ▶ If RVs are independent \Rightarrow Variance of sum is sum of variances
- ▶ Slightly more general result holds for independent X_i , $i = 1, \dots, n$

$$\text{var}\left[\sum_i (a_i X_i + b_i)\right] = \sum_i a_i^2 \text{var}[X_i]$$

Variance of binomial RV and sample mean

Ex: Let $X_i, i = 1, \dots, n$ be independent Bernoulli RVs

\Rightarrow Recall $\mathbb{E}[X_i] = p$ and $\text{var}[X_i] = p(1 - p)$

► Write **binomial** X with parameters (n, p) as: $X = \sum_{i=1}^n X_i$

► Variance of binomial then $\Rightarrow \text{var}[X] = \sum_{i=1}^n \text{var}[X_i] = np(1 - p)$

Ex: Let $Y_i, i = 1, \dots, n$ be independent RVs and $\mathbb{E}[Y_i] = \mu$, $\text{var}[Y_i] = \sigma^2$

► **Sample mean** is $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. What about $\mathbb{E}[\bar{Y}]$ and $\text{var}[\bar{Y}]$?

► Expected value $\Rightarrow \mathbb{E}[\bar{Y}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] = \mu$

► Variance $\Rightarrow \text{var}[\bar{Y}] = \frac{1}{n^2} \sum_{i=1}^n \text{var}[Y_i] = \frac{\sigma^2}{n}$ (used independence)

- ▶ **Def:** The **covariance of X and Y** is (generalizes variance to pairs of RVs)

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- ▶ If $\text{cov}(X, Y) = 0$ variables X and Y are said to be **uncorrelated**
- ▶ If X, Y independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $\text{cov}(X, Y) = 0$
⇒ **Independence implies uncorrelated RVs**
- ▶ Opposite is **not** true, may have $\text{cov}(X, Y) = 0$ for dependent X, Y
 - ▶ **Ex:** X uniform in $[-a, a]$ and $Y = X^2$
⇒ **But uncorrelatedness implies independence if X, Y are normal**
- ▶ If $\text{cov}(X, Y) > 0$ then X and Y tend to move in the same direction
⇒ **Positive correlation**
- ▶ If $\text{cov}(X, Y) < 0$ then X and Y tend to move in opposite directions
⇒ **Negative correlation**

- ▶ Let X be a zero-mean random signal and Z zero-mean noise
 \Rightarrow Signal X and noise Z are independent
- ▶ Consider received signals $Y_1 = X + Z$ and $Y_2 = -X + Z$
- (I) Y_1 and X are **positively correlated** (X, Y_1 move in same direction)

$$\begin{aligned}\text{cov}(X, Y_1) &= \mathbb{E}[XY_1] - \mathbb{E}[X]\mathbb{E}[Y_1] \\ &= \mathbb{E}[X(X + Z)] - \mathbb{E}[X]\mathbb{E}[X + Z]\end{aligned}$$

- ▶ Second term is 0 ($\mathbb{E}[X] = 0$). For first term independence of X, Z

$$\mathbb{E}[X(X + Z)] = \mathbb{E}[X^2] + \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[X^2]$$

- ▶ Combining observations \Rightarrow **$\text{cov}(X, Y_1) = \mathbb{E}[X^2] > 0$**

(II) Y_2 and X are **negatively correlated** (X , Y_2 **move opposite direction**)

- ▶ Same computations $\Rightarrow \text{cov}(X, Y_2) = -\mathbb{E}[X^2] < 0$

(III) Can also compute correlation between Y_1 and Y_2

$$\begin{aligned}\text{cov}(Y_1, Y_2) &= \mathbb{E}[(X + Z)(-X + Z)] - \mathbb{E}[(X + Z)] \mathbb{E}[(-X + Z)] \\ &= -\mathbb{E}[X^2] + \mathbb{E}[Z^2]\end{aligned}$$

\Rightarrow Negative correlation if $\mathbb{E}[X^2] > \mathbb{E}[Z^2]$ (small noise)

\Rightarrow Positive correlation if $\mathbb{E}[X^2] < \mathbb{E}[Z^2]$ (large noise)

- ▶ Correlation between X and Y_1 or X and Y_2 comes from causality
- ▶ Correlation between Y_1 and Y_2 does not. **Latent variables X and Z**
 \Rightarrow **Correlation does not imply causation**

Plausible, indeed commonly used, model of a communication channel

- ▶ Sample space
- ▶ Outcome and event
- ▶ Sigma-algebra
- ▶ Countable union
- ▶ Axioms of probability
- ▶ Probability space
- ▶ Conditional probability
- ▶ Law of total probability
- ▶ Bayes' rule
- ▶ Independent events
- ▶ Random variable (RV)
- ▶ Discrete RV
- ▶ Bernoulli, binomial, Poisson
- ▶ Continuous RV
- ▶ Uniform, Normal, exponential
- ▶ Indicator RV
- ▶ Pmf, pdf and cdf
- ▶ Law of rare events
- ▶ Expected value
- ▶ Variance and standard deviation
- ▶ Joint probability distribution
- ▶ Marginal distribution
- ▶ Random vector
- ▶ Independent RVs
- ▶ Covariance
- ▶ Uncorrelated RVs